SWE2020- SOFTWARE METRICS
FINAL REVIEW-3

Name: E.SravenKumar

Regno.: 16MIS0482

Slot: B1

Data Set Name: **Autistic Spectrum Disorder Screening Data for Adult**

URL of the dataset: **https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult**

Attribute information:

In this dataset there are 21 attributes and 704 instances. Attributes are

```
@attribute A1_Score {0,1}
@attribute A2_Score {0,1}
@attribute A3_Score {0,1}
@attribute A4_Score {0,1}
@attribute A5_Score {0,1}
@attribute A6_Score {0,1}
@attribute A7_Score {0,1}
@attribute A8_Score {0,1}
@attribute A9_Score {0,1}
@attribute A10_Score {0,1}
@attribute age numeric
@attribute gender {f,m}
@attribute ethnicity
@attribute jundice {no,yes}
@attribute austim {no,yes}
@attribute contry_of_res
@attribute used_app_before {no,yes}
@attribute result numeric
@attribute age_desc {'18 and more'}
@attribute relation {Self,Parent,'Health care
professional',Relative,Others}
@attribute Class/ASD {NO,YES}
```

Classification model analysis:

For this dataset it is possible to evaluate the models like decision tree, SVM, and neural network. To find out which model is more suitable for our dataset we considered error rate as an important attribute the following are the results:

Model	Error rate
Decision tree	O%
SVM	1.85%
Neural network	4.52%

By considering all the inputs we can get decision tree is the best model for this data set. Because it gives low error rate than other models.

Summary:

To check missing tolerance level for the models we eliminated two attributes they are scoure and

jaundies. And the result is

Model	Error rate
Decision tree	O%
SVM	3.7%
Neural network	3.15%

As case 2 we eliminated other two attributes gender and ethnicity and the result is

Model	Error rate
Decision tree	O%
SVM	1.85%
Neural network	1.6%

As case 3 we eliminated first five attributes and as result we get

Model	Error rate
Decision tree	0%
SVM	0%
Neural network	50%

No.of leaf nodes=6

	Decision tree	Svm	linear	Neural net.	Best model
Ds1	0	0	60	0	DT,SVM,NN
Ds2	25	25	0	0	Linear,NN
Ds3	27.7	33.3	22.2	27.8	Linear
Ds4	25	25	75	0	NN
Ds5	57	28	21	21	Linear,NN
Ds6	67	50	28	39	Linear
Total dataset	34.3	29	20.3	23.8	Linear

Cumulative error value with clustering: 3.86

Error rate without clustering: 4.52

Reduced error value =Cumulative error value with clustering - Error rate without clustering

➔ 4.52-3.86 =0.66 %

Conclusion:

0.66% Error rate is reduced after clustering.